IBM Applied Data Science Capstone Project by Coursera

Opening a new Coffee Café in Bengaluru, India

Girish A Koushik

June 2020

Introduction

"When you live in South India, coffee is your lifeline". This is how a regional newspaper (Bangalore Mirror) article describes the coffee consumption in Bengaluru, which is the largest city in South India, the capital of the southern state of Karnataka and the 'Silicon Valley' of the Indian sub-continent. The article goes on quoting consumers who would sometimes love to consume more than 5 cups of coffee a day, without worrying much about the medical research that presents the demerits of very high consumption of caffeine. The southern region of India accounts for the largest share of coffee in the beverage market with 73 percent.

The coffee shops in Bengaluru are frequented by customers who cherish either the 'traditional filter coffee' or the cappuccino, espresso and other 'international variants' of coffee. The most recent trend in the city's coffee culture stems from the growing emphasis on artisan coffee and the perception of coffee as an experience rather than just a daily dose of caffeine. This comes from an increasing interest in quality coffee among the millennial who look at coffee consumption as an aspirational experience, a new lifestyle, especially as far as out-of-home coffee consumption.

In 2020, the number of organized café chains across India was forecast to increase to around 6,200, up from about 3,500 in 2017. The value of coffee in the retail market across India was forecast to be 140 Billion Indian Rupees in 2020. To cater to the demand of the working population, new coffee shops are being opened across the city and hence, the location of the coffee café is one of the most important decisions that will determine whether the café will be a success or not.

Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Bengaluru, India to open a new coffee café. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Bengaluru, if a business owner is looking to open a new coffee café, where would you recommend that they open it?

Target Audience of the Project

The findings of this project are not only useful to small business owners who are looking to open a new coffee café and/or expand their business across their city but also to popular coffee chains such as Starbucks, Barista, Café Coffee Day etc., who are looking to increase their customer base.

Data Requirements

To solve the problem, we require the following data:

1. List of neighborhoods in Bengaluru. This defines the scope of the project, which is confined to the city of Bengaluru (capital of the state of Karnataka) in India.
2. Latitude and Longitude of the above neighborhoods. This is required in order to obtain the venue data and to plot the map.
3. Venue data, particularly related to the coffee shops. We will use this data to perform the clustering on the neighborhoods.

Data Collection

The Wikipedia page (Category: Neighborhoods in Bangalore) contains the list of neighborhoods in Bengaluru having a total of 128 neighborhoods. Web scraping techniques will be used to extract the data from the Wikipedia page, with the help of Python requests and 'beautiful soup' packages. The python 'Geocoder' package will retrieve the latitude and longitude coordinates of the neighbourhoods.

Further, the Foursquare API will be used to get the venue data for the above neighbourhoods. Foursquare API will provide many categories of the venue data and the project interest will be the 'Coffee Shops' category in order to solve the business problem. This project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (k - means clustering) and map visualization (Folium).